An Enhancement approach for AI based early forest fire detection using KNN algorithm

Ramya.S
M.E., Department of ECE,
Kongunadu College of Engineering and Technology,
Trichy,India.

Krishnakumar.V
Assistant Professor, Department of ECE,
Kongunadu College of Engineering and Technology,
Trichy,India.

Abstract- Forests are the protectors of earth's ecological balance. Unfortunately, the forest fire is usually only observed when it has already spread over a large area, making its control land stoppage arduous and even impossible at times. The result is devastating loss and irreparable damage to the environment and atmosphere. Fast and effective detection is a key factor in forest fire fighting. To avoid such wide spreading of forest fires it is necessary to detect fires in an early state and to prevent the propagation. Every year, thousands of forest fire across the globe cause disasters beyond measure and description. Currently, there are many different solutions to detect the forest fires. People are using sensors to detect the fire. But this case is not possible for large acres of forest. The existing system is based on tracking and alarming for the protection of trees against forest fires. In this paper, we discuss a new approach for fire detection, in which modern technologies are used. We propose a platform based on Artificial Intelligence. Machine learning with KNN algorithm is used here for finding the amount of smoke and fire.

I .INTRODUCTION

Every year, thousands of forest fire across the globe cause disasters beyond measure and description. If a fire is clearly identified, fire suppression is initialized by an alarm going directly to the fire brigade. Nowadays, Wireless Sensor Networks (WSNs) are critical components of the increasingly common IoT (Internet of Things) systems. Such systems have large applicability, and the environmental monitoring field can also benefit from their innovation. The purpose of the IoT concept is to transform the real world and every day electronic devices, appliances, etc., into intelligent interconnected virtual objects. By keeping the user informed on the state of things and giving the users control of things, a better global humans-devices-humans communication can be achieved. Artificial intelligence (AI), sometimes called machine intelligence, is the intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans. Artificial intelligence (AI) is an area of computer science that emphasizes the creation of intelligent machines that work and react like humans. The main objective is to detect the presence of fire and smoke using artificial intelligence method.

II. CONCEPTUAL MODEL FOR EARLY FOREST FIRE DETECTION USING MACHINE LEARNING TECHNIQUE

This section describes the conceptual model of a platform for early forest fire detection that is going to be implemented using KNN algorithm.

A. *Dataset*

A dataset is a collection of data.

Training dataset

A dataset that is fed into our machine learning algorithm is to train the model. A training dataset is a dataset of examples used for learning that is to fit the parameters.

Testing dataset

A dataset which is used to validate the accuracy of model but is not used to train the model. A test dataset is a dataset that is distribution as the training dataset. Independent of the training dataset, but that follows the same probability.

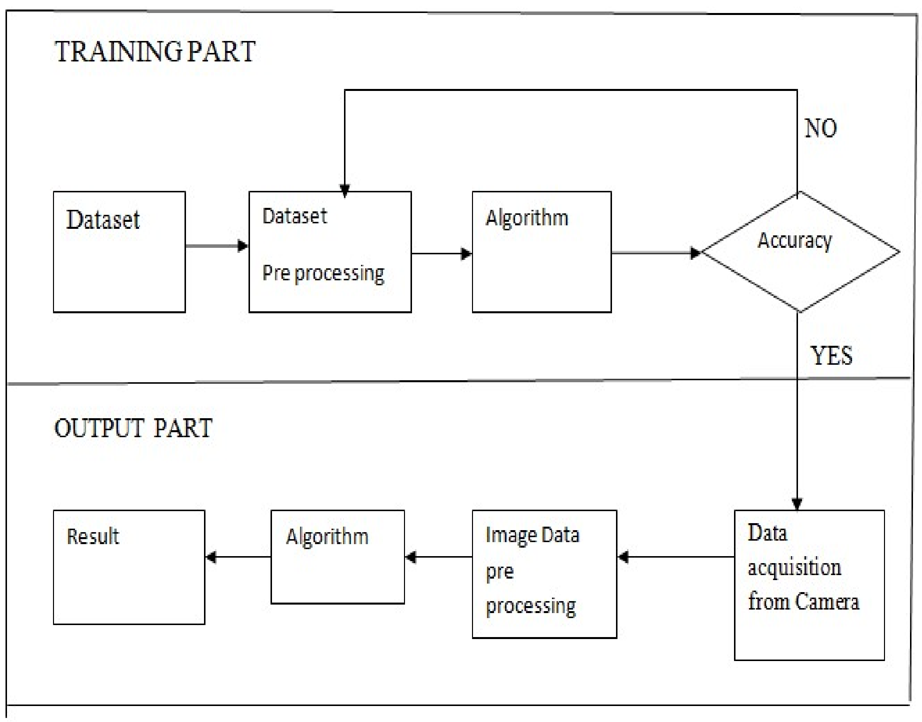
B. THE DIFFERENT DATA SETS OF MACHINE LEARNING

Machine learning typically works with two data sets: training and test. All should randomly sample a larger body of data. The first set is the training set, the largest of the three. Running a training set through a neural network teaches the net how to weigh different features, adjusting them coefficients according to their likelihood of minimizing errors in your results. Those coefficients, also known as parameters, will be contained in tensors and together they are called the model because they encode a model of the data which is trained on. Thus the most important takeaways are obtained from training a neural network.

The second set is test set. It functions as a seal of approval, and it is tested against final random sampling. The result it produces should validate and recognizes images.

In case of no accurate predictions, it is moved back to the training set. Hence look at the hyper parameters used to tune the network again, as well as the quality of data and look at the pre-processing techniques.

III . BLOCK DIAGRAM



IV. PREPROCESSING

(i) TEST_TRAIN-SPLIT

Here this involves the separation of variables into dependent and independent where it undergoes Test_Train_split which are separated into four sets as X-train, Xtest, Y-train and Y-test. Thus the trained datas are separated as 750 and the test datas as 250 (X-train=750, X-test=250, Y-test=250 and Y-train=750).

Finally the trained datas X-train and Y-train are given into the KNN algorithm and are trained. Thus after the training when the test datas of X-test is given, it gives a predicted value with accuracy using confusion matrix.

Thus if there is less accuracy than 90% it will be taken for Preprocessing again.

On the other the hand the predicted values is compared with the remaining Y- test and accuracy is obtained which lead to the trained module. Thus the new data can be given now in which the data can be preprocessed. The preprocessing includes Resize, Blur and HSV.

(ii) RESIZE

Image resizing refers to scaling of images. Scaling comes handy in many image processing as well as machine learning applications. It also helps in zooming in images. Many times resizing the image involves i.e. either shirk it or scale up to meet the size requirements. An image loses quality when zoom is applied. When a small quantity of pixels is put onto the screen with higher resolution, it is necessary to "create" new pixels, so that can able to occupy the holes that would appear.

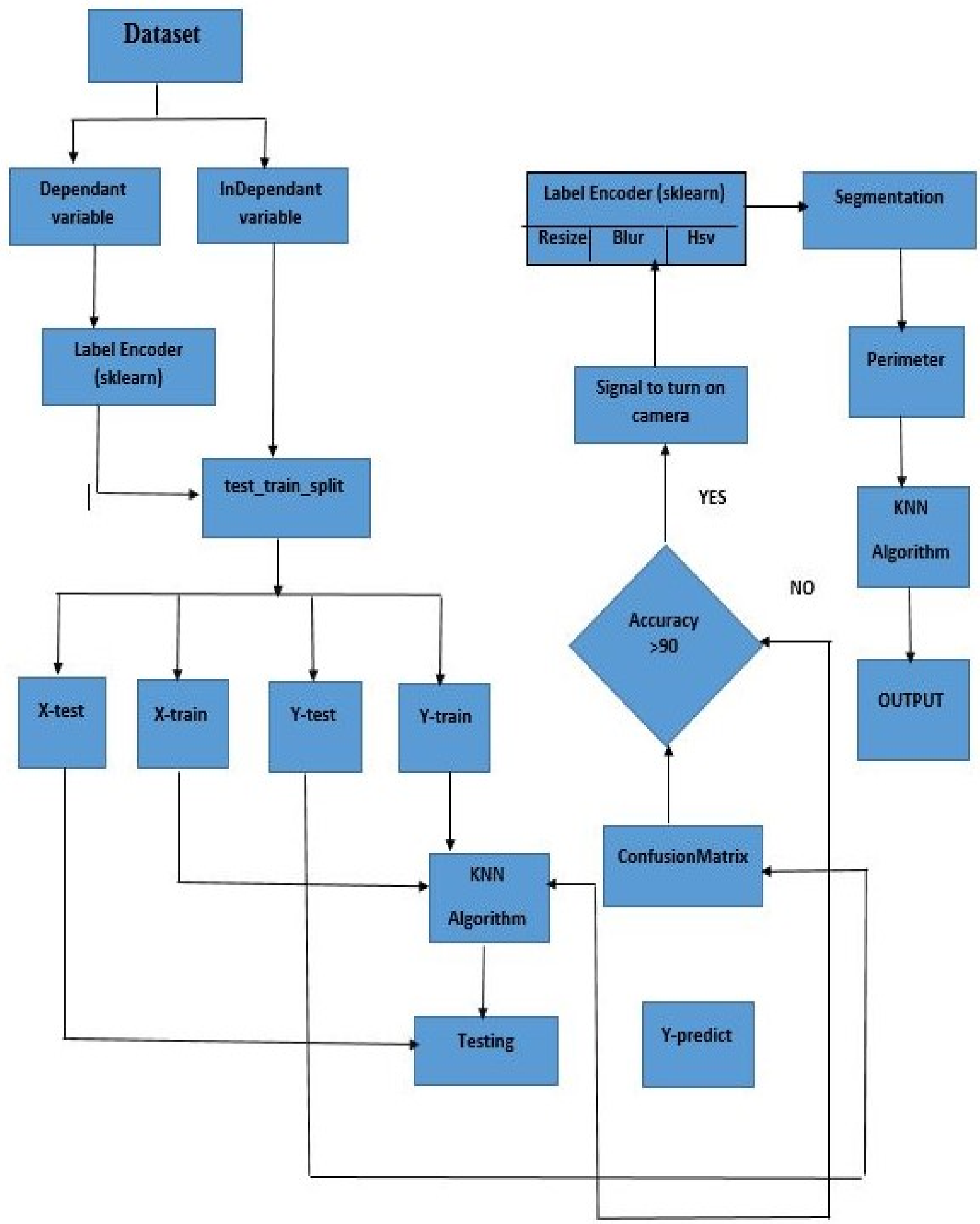
(iii) BLUR

An image looks sharper or more detailed able to perceive all the objects and shapes correctly in it. This shape of an object is due to its edges. So in blurring, the edge contents are reduced and make the transition from one color to the other very smooth. “Gaussian filter” is used here to blur the image.

(iv) HSV

There are various models one of which is the hue, saturation, value (HSV) model. Using this model, an object with a certain color can be detected and to reduce the influence of light intensity from the outside. Thus finally the data is segmented where color segmentation takes place here. Then the perimeter is measured and the predicted data is obtained as output here

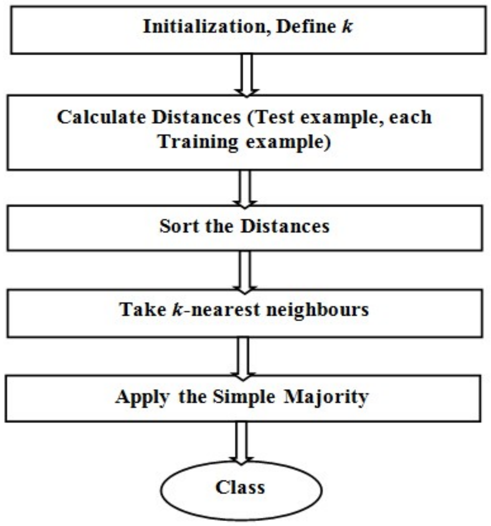
V. PROPOSED FLOW CHART



VI.ALGORITHM USED

It's a simple algorithm that stores all available cases and classifies any new cases by taking a majority vote of its k neighbors. K-Nearest Neighbors (KNN) is one of the simplest algorithms used in machine learning for regression and classification problem. When the training data is given into an algorithm, then this algorithm uses the training data to give predictions on a new test data.

KNN FLOWCHART



Thus a test data is taken, and finds k nearest data values to this from the data set. Then it finally selects the neighbor of maximum frequency.

This algorithm can be applied to both classification and regression problems. Apparently, within the Data Science industry, it's more widely used to solve classification problems.

It’s a simple algorithm that stores all available cases and classifies any new cases by taking a majority vote of its k neighbors. The case is then assigned to the class with which it has the most in common. A distance function performs this measurement.

• KNN is computationally expensive .Variables should be normalized, or else higher range variables can bias the algorithm.

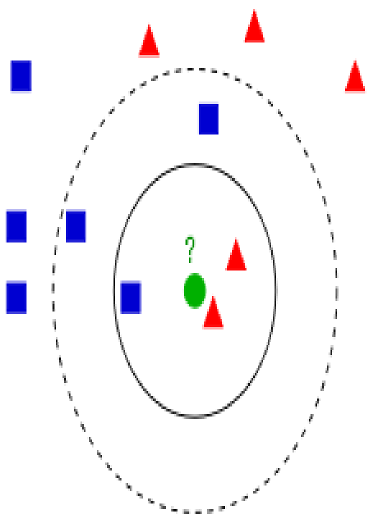
• The k-Nearest-Neighbor Classifier (k-NN) works directly on the learned samples, instead of creating rules compared to other classification methods.

CHOOSING THE RIGHT VALUE FOR ‘K’

To select the K that’s right for required data, there is a need to run the KNN algorithm several times with different values of K and choose the K that reduces the number of errors during encounter while maintaining the algorithm’s ability is to accurately make predictions when the given data it not seen before.

The principle behind nearest neighbour classification consists in finding a predefined number, i.e. the 'k' - of training samples closest in distance to a new sample, which has to be classified. The label of the new sample will be defined from these neighbours. K-nearest neighbour classifiers have a fixed user defined constant for the number of neighbor which have to be determined. There are also radius based neighbour learning algorithms, which have a varying number of neighbours based on the local density of points, all the samples inside of a fixed radius. The distance can, in general, be any metric measure: standard Euclidean distance is the most common choice. Neighbours-based methods are known as non-generalizing machine learning methods, since simply "remember" all of its training data. Classification can be computed by a majority vote of the nearest neighbours of the unknown sample.

KNN EXAMPLE DIAGRAM



From the above diagram, the test sample (green dot) should be classified either to blue squares or to red triangles. If k = 3 (solid line circle) it is assigned to the red triangles because there are 2 triangles and only 1 square inside the inner circle. If k = 5 (dashed line circle) it is assigned to the blue squares (3 squares vs. 2 triangles inside the outer circle).

ADVANTAGES OF KNN ALGORITHM

The algorithm is simple and easy to implement. There's no need to build a model, tune several parameters, or make additional assumptions. Robust with regard to the search space; for instance, classes don't have to be linearly separable.

LABEL ENCODING

Label Encoding refers to converting the labels into numeric form so as to convert it into the machine-readable form. Label Encoding in Python can be achieved using Sklearn Library. Sklearn provides a very efficient tool for encoding the levels of categorical features into numeric values. Label Encoder encodes labels with a value between 0 and n_classes-1 where n is the number of distinct labels. If a label repeats it assigns the same value to as assigned earlier.

original dataset

x_1	x_2	x_3	y
5	efik	8	calabar
9	ibibio	3	uyo
8	igbo	6	owerri
0	ibibio	5	uyo
2	efik	3	calabar
0	efik	8	calabar
1	igbo	8	owerri

LabelEncoder →

{
"efik" --> 0
"ibibio" --> 1
"igbo" --> 2
}

encode categorical feature

x_1	x_2	x_3	y
5	0	8	calabar
9	1	3	uyo
8	2	6	owerri
0	1	5	uyo
2	0	3	calabar
0	0	8	calabar
1	2	8	owerri

Thus the SciKit Learn library in Python is used to convert categorical data, or text data, into numbers, which our predictive models can better understand.

CONFUSION MATRIX

A confusion matrix is a summary of prediction results on a classification problem. The number of correct and incorrect predictions are summarized with count values and broken down by each class. This is the key to the confusion matrix. The confusion matrix shows the ways in which the classification model is confused when it makes predictions. It gives insight not only into the errors being made by a classifier but more importantly the types of errors that are being made.

	Class 1 Predicted	Class 2 Predicted
Class 1 Actual	TP	FN
Class 2 Actual	FP	TN

Here

•Class1:Positive

• Class 2: Negative

DEFINITION OF THE TERMS

• Positive (P): Observation is positive (for example: is an apple).

• Negative (N): Observation is not positive (for example: is not an apple).

• True Positive (TP): Observation is positive, and is predicted to be positive.

• False Negative (FN): Observation is positive, but is predicted negative.

• True Negative (TN): Observation is negative, and is predicted to be negative.

• False Positive (FP): Observation is negative, but is predicted positive.

CLASSIFICATION RATE/ACCURACY

For binary classification, classification rate or accuracy can also be calculated in terms of positives and negatives as follows:

Accuracy =TP+TN/ TP+TN+FP+FN

Where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

However, there are problems with accuracy. It assumes equal costs for both kinds of errors. Finally 99% accuracy can be excellent, good, mediocre, poor or terrible depending upon the problem.

RECALL

Recall can be defined as the ratio of the total number of correctly classified positive examples divide to the total number of positive examples. High Recall indicates the class is correctly recognized (small number of FN).Recall is given by the relation as follows:

Recall =TP/ TP+FN

PRECISION

To get the value of precision divide the total number of correctly classified positive examples by the total number of predicted positive examples. High Precision indicates an example labeled as positive is indeed positive (small number of FP). Precision is given by the relation:

Precision=TP/ TP+FN

(i)HIGH RECALL, LOW PRECISION

This means that most of the positive examples are correctly recognized (low FN) but there are a lot of false positives.

(ii)LOW RECALL, HIGH PRECISION

This shows that there is a lot of positive examples (high FN) but those we predict as positive are indeed positive (low FP)

VIII.SOFTWARE DESCRIPTION

PYTHON

Python is a wonderful and powerful programming language that's easy to use (easy to read and write) and with Raspberry Pi lets you connect your project to the real world. Python is an interpreter, high-level, general purposes programming language. Python’s design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

Python syntax is very clean, with an emphasis on readability and uses Standard English keywords. It is started by opening IDLE from the desktop. Python interpreters are available for many operating systems.

PYTHON IN ARTIFICIAL INTELLIGENCE

Python combines remarkable power with very clear syntax. Python is also used as an extension language for applications written in other languages that need easy-to-use scripting or automation interfaces. The language is great to use when working with machine learning algorithms and has easy syntax relatively

. VIII.RESULTS

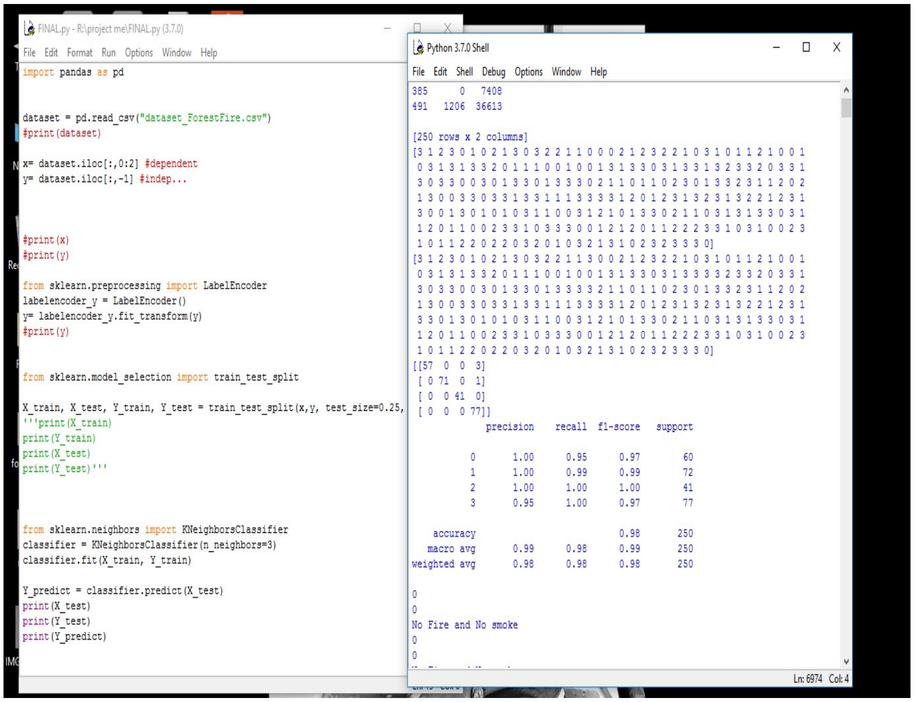


Figure 1 :O/P python shell representing accuracy

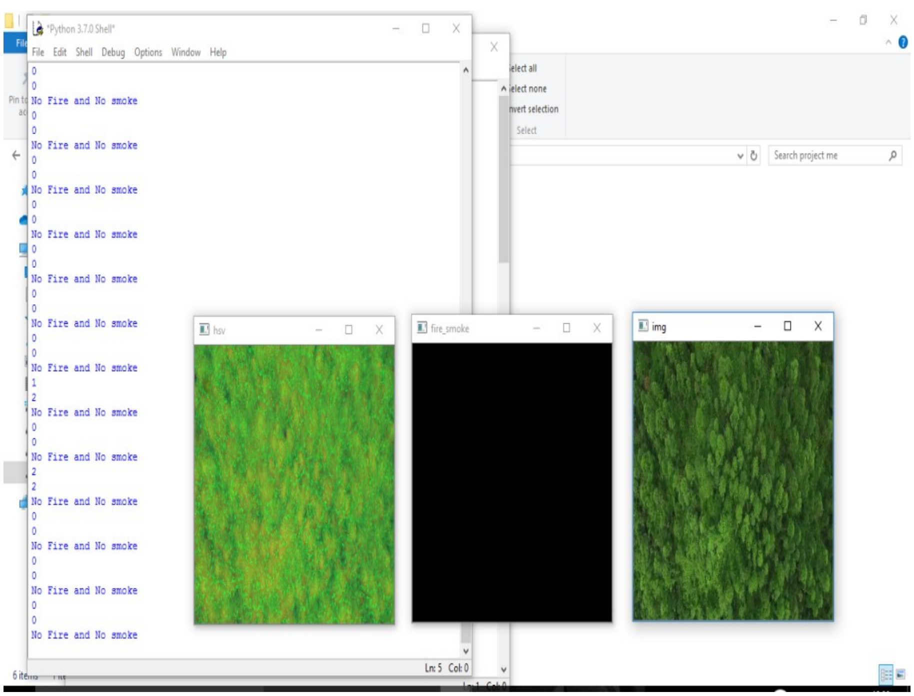


Figure 2 : O/P window representing No fire, no smoke

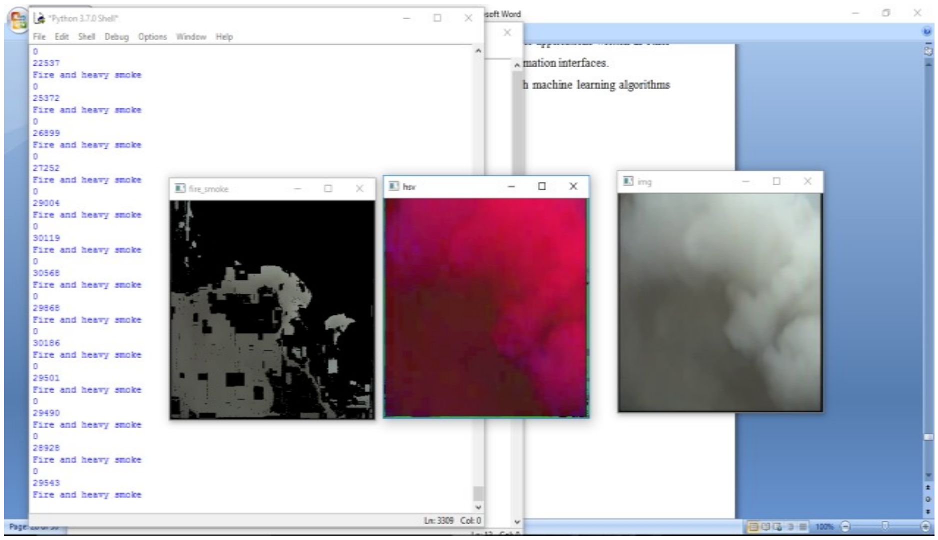


Figure 3: O/p window representing Fire and heavy smoke

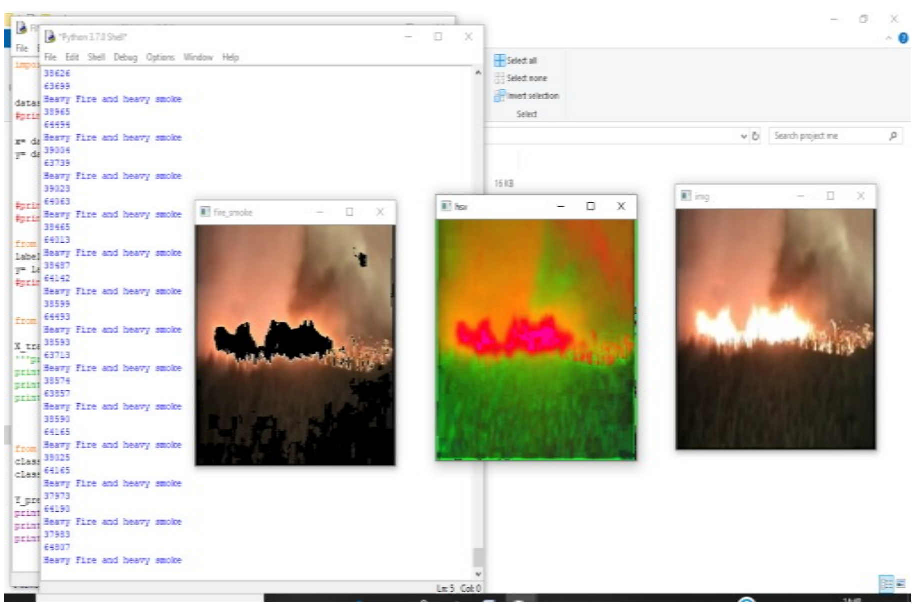


Figure 4: o/p window representing heavy fire and heavy smoke

IX.CONCLUSION

An active machine learning (KNN) approach is proposed for the Forest fire detection with the greater automation and better performance. The development approach was evaluated with actual datasets collected from the images while capturing the forest fire. The evaluation process is conducted with manually labelled data and the proposed active machine learning shows a favourable performance. The accuracy of forest fire prediction is to be 98% using KNN algorithm. From this better performance analysis is obtained.

REFERENCES

[1] Ambrosia.V.G , S. Wegener, T. Zajkowski, D. V. Sullivan, S. Buechel, ”The Ikhana unmanned airborne system (UAS) western states fire imaging missions: from concept to reality (20062010)”, Geocarto International, vol. 26, no. 2, 2011, pp. 85– 101.

[2] A. Filonenko and K. H. Jo, ”Fast fire flame detection on videos using AdaBoost and parallel processing”, Proceedings of the 2018 IEEE Industrial Cyber-Physical Systems (ICPS), 2018, pp. 645–650.

[3] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, “Internet of things: A survey on enabling technologies, protocols, and applications,” IEEE Communications Surveys and Tutorials, vol. 17, no. 4, pp. 2347–2376, 2015.

[4] Chen, Thou-Ho, Cheng-Liang Kao, and Sju-Mo Chang. "An intelligent real-time fire-detection method based on video processing." Security Technology, 2003. Proceedings. IEEE 37th Annual 2003 International Carnahan Conference on. IEEE, 2003.

[5] Christensen.B R , ”Use of UAV or remotely piloted aircraft and forward-looking infrared in forest, rural and wildland fire management: evaluation using simple economic analysis”, New Zealand Journal of Forestry Science, vol. 45, no. 1, 2015, p. 16.

[6] Chen, Thou-Ho, et al. "The smoke detection for early firealarming system base on video processing." Intelligent Information Hiding and Multimedia Signal Processing, 2006. IIHMSP'06. International Conference on. IEEE, 2006.

[7] Cleetus.N , and K. A. Dhanya, “Multi-objective functions in particle swarm optimization for intrusion detection,” in International Conference on Advances in Computing, Communications and Informatics, 2014, pp. 387–392.

[8] Farah.N , M. Avishek, F. Muhammad, A. Rahman, M. Rafni, and D. Md, “Application of machine learning approaches in intrusion detection system: A survey,” International Journal of Advanced Research in Artificial Intelligence, vol. 4, no. 3, 2015.

[9] Ghamry.K.A , M. A. Kamel and Y. Zhang, ”Cooperative forest monitoring and fire detection using a team of UAVs-UGVs”, Proceedings of 2016 International Conference on Unmanned Aircraft Systems (ICUAS), 2016, pp.1206–1211.

[10] Kakiz.M.T ,E.Ozturk,andT.Cavdar,“AnovelSDN-basedIoT architecture for big data,” in International Artificial Intelligence and Data Processing Symposium, 2006, pp. 1–5.

[11] Kanungo.T , D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, “An efficient k-means clustering algorithm: analysis and implementation,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 881–892, 2002

[12] Liau.H , Y. Nimmagadda and Y. Wong, ”Fire SSD: Wide fire modules based single shot detector on edge device”, Web: http://arxiv.org/abs/1806.05363.

[13] Mahdavinejad.M.S , M. Rezvan, M. Barekatain, P. Adibi, P. Barnaghi, and A. P. Sheth, “Machine learning for Internet of Things data analysis: A survey,” Digital Communications and Networks, 2017.

[14] Noda.S , and K. Ueda. "Fire detection in tunnels using an image processing method." Vehicle Navigation and Information Systems Conference, 1994. Proceedings., 1994. IEEE, 1994.

[15] Ofli.F , P. Meier, M. Imran, C. Castillo, D. Tuia, ”Combining human computing and machine learning to make sense of big (aerial) data for disaster response”, Big Data, vol. 4, no. 1, 2016, pp. 47–59.

[16] Rao.G.N , P. J. Rao and R. Duvvuru, “A drone remote sensing for virtual reality simulation system for forest fires: semantic neural network approach”, Proceedings of IOP Conference Series: Materials Science and Engineering, vol. 149, no. 1, 2016, p. 12011.

[17] Ding, Y. Chen, Z. Wang, and H. Zhang, “Intelligent 5G: When cellular networks meet artificial intelligence,” IEEE Wireless Communications, vol. PP, no. 99, pp. 2–10, 2017.

[18] Salman.O , S. Abdallah, I. H. Elhajj, A. Chehab, and A. Kayssi, “Identity-based authentication scheme for the internet of things,” Computers and Communication, pp. 1109–1111, 2016.

[19] Shen.D, X. Chen, M. Nguyen and W. Q. Yan, “Flame detection using deep learning”, Proceedings of the IEEE 4th International Conference on Control, Automation and Robotics (ICCAR), 2018, pp. 416–420.

[20] Sherstjuk.V, Zharikova.M , and Sokol.I, “Forest Fire-Fighting Monitoring System Based on UAV team and Remote Sensing”, Proceedings of 2018 IEEE 38th International Conference on Electronics and Nanotechnology (ELNANO), 2018, pp. 663–6.

[21] Tang.A.T, L. Mhamdi, D. Mclernon, S. A. R. Zaidi, and M. Ghogho, “Deep learning approach for network intrusion detection in software defined networking,” in International Conference on Wireless Networks and Mobile Communications, 2018.

[22] Wang, Da-Jinn, Yen-Hui Yin, and Tsong-Yi Chen. "Smoke Detection for Early Fire-Alarming System Based on Video Processing." Journal of Digital Information Management 6.2 (2008).

[23] Yu.L, N. Wang and X. Meng, ”Real-time forest fire detection withwireless sensor networks”, Proceedings of 2005 International Conference on Wireless Communications, Networking and Mobile Computing, vol. 2, 2005, pp. 1214–1217.

[24] Yuan.C, Z. Liu and Y. Zhang, “UAV-based forest fire detection and tracking using image processing techniques”, Proceedings of 2015 International Conference on Unmanned Aircraft Systems (ICUAS), 2015, pp. 639–643.

[25] Zhang.Q, G. Lin, Y. Zhang, G. Xu and J. Wang, “Wildland forest fire smoke detection based on faster R-CNN using synthetic smoke images”, Procedia Engineering, vol. 211, 2018, pp. 441–446.